

TITLE

## PLANT AMINO ACID BIOSYNTHETIC ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/048,771, filed June 6, 1997, and U.S. Provisional Application No. 60/049,443, filed June 12, 1997.

5

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding enzymes involved in amino acid biosynthesis in plants and seeds.

BACKGROUND OF THE INVENTION

10

Many vertebrates, including man, lack the ability to manufacture a number of amino acids and therefore require these amino acids preformed in the diet. These are called essential amino acids. Human food and animal feed, derived from many grains, are deficient in essential amino acids, such as lysine, the sulfur amino acids methionine and cysteine, threonine and tryptophan. For example, in corn (*Zea mays L.*) lysine is the most limiting amino acid for the dietary requirements of many animals. Soybean (*Glycine max L.*) meal is used as an additive to corn-based animal feeds primarily as a lysine supplement. Thus, an increase in the lysine content of either corn or soybean would reduce or eliminate the need to supplement mixed grain feeds with lysine produced via fermentation of microbes.

15

Furthermore, in corn the sulfur amino acids are the third most limiting amino acids, after lysine and tryptophan, for the dietary requirements of many animals. The use of soybean meal, which is rich in lysine and tryptophan, to supplement corn in animal feed is limited by the low sulfur amino acid content of the legume. Thus, an increase in the sulfur amino acid content of either corn or soybean would improve the nutritional quality of the mixtures and reduce the need for further supplementation through addition of more expensive methionine.

20

Lysine, threonine, methionine, cysteine and isoleucine are amino acids derived from aspartate. Regulation of the biosynthesis of each member of this family is interconnected (see Figure 1). One approach to increasing the nutritional quality of human foods and animal feed is to increase the production and accumulation of specific free amino acids via genetic engineering of this biosynthetic pathway. Alteration of the activity of enzymes in this pathway could lead to altered levels of lysine, threonine, methionine, cysteine and isoleucine. However, few of the genes encoding enzymes that regulate this pathway in plants, especially corn, soybeans and wheat, are available.

25

30

The organization of the pathway leading to biosynthesis of lysine, threonine, methionine, cysteine and isoleucine indicates that over-expression or reduction of expression of genes encoding, *inter alia*, threonine synthase, dihydrodipicolinate reductase, diaminopimelate epimerase, threonine deaminase and S-adenosylmethionine synthetase in corn, soybean, wheat and other crop plants could be used to alter levels of these amino acids in human food and animal feed. Accordingly, availability of nucleic acid sequences

35

encoding all or a portion of these enzymes would facilitate development of nutritionally improved crop plants.

### SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding plant enzymes involved in amino acid biosynthesis. Specifically, this invention concerns isolated nucleic acid fragments encoding the following plant enzymes that catalyze steps in the biosynthesis of lysine, threonine, methionine, cysteine and isoleucine from aspartate: dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase and S-adenosylmethionine synthetase. In addition, this invention relates to nucleic acid fragments that are complementary to nucleic acid fragments encoding the listed plant biosynthetic enzymes.

In another embodiment, the instant invention relates to chimeric genes encoding the amino acid biosynthetic acid enzymes listed above or to chimeric genes that comprise nucleic acid fragments that are complementary to the nucleic acid fragments encoding the enzymes, operably linked to suitable regulatory sequences, wherein expression of the chimeric genes results in production of levels of the encoded enzymes in transformed host cells that are altered (i.e., increased or decreased) from the levels produced in untransformed host cells.

In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding a plant amino acid biosynthetic enzyme operably linked to suitable regulatory sequences, the enzyme selected from the group consisting of: dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase and S-adenosylmethionine synthetase. Expression of the chimeric gene results in production of altered levels of the biosynthetic enzyme in the transformed host cell. The transformed host cells can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

An additional embodiment of the instant invention concerns a method of altering the level of expression of a plant biosynthetic enzyme in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a plant biosynthetic enzyme selected from the group consisting of dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase and S-adenosylmethionine synthetase, operably linked to suitable regulatory sequences; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of the biosynthetic enzyme in the transformed host cell.

An additional embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or substantially all of an amino acid sequence encoding a

plant dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase and S-adenosylmethionine synthetase.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of a plant biosynthetic enzyme selected from the group consisting of dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase and S-adenosylmethionine synthetase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a plant biosynthetic enzyme selected from the group consisting of dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase and S-adenosylmethionine synthetase, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the biosynthetic enzyme in the transformed host cell; (c) optionally purifying the biosynthetic enzyme expressed by the transformed host cell; (d) treating the biosynthetic enzyme with a compound to be tested; and (e) comparing the activity of the biosynthetic enzyme that has been treated with a test compound to the activity of an untreated biosynthetic enzyme, thereby selecting compounds with potential for inhibitory activity.

#### BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and sequence descriptions which form a part of this application.

Figure 1 depicts the biosynthetic pathway for the aspartate family of amino acids. The following abbreviations are used: AK = aspartokinase; ASADH = aspartic semialdehyde dehydrogenase; DHDPS = dihydrodipicolinate synthase; DHDPR = dihydrodipicolinate reductase; DAPEP = diaminopimelate epimerase; DAPDC = diaminopimelate decarboxylase; HDH = homoserine dehydrogenase; HK = homoserine kinase; TS = threonine synthase; TD = threonine deaminase; CyS = cystathionine  $\gamma$ -synthase; C $\beta$ L = cystathionine  $\beta$ -lyase; MS = methionine synthase; CS = cysteine synthase; and SAMS = S-adenosylmethionine synthase.

Figure 2 shows a multiple alignment of the amino acid sequence fragments reported herein encoding dihydrodipicolinate reductase (SEQ ID NOs:2 and 4) and the *Synechocystis* sp. dihydrodipicolinate reductase sequence set forth in DDBJ Accession No. D90899 (SEQ ID NO:5).

Figure 3 shows a multiple alignment of the amino acid sequence fragments reported herein encoding diaminopimelate epimerase (SEQ ID NOs:7, 9, 11, and 13) and the *Synechocystis* sp. diaminopimelate epimerase sequence set forth in DDBJ Accession No. D90917 (SEQ ID NO:14).

Figure 4 shows a multiple alignment of the amino acid sequence fragments reported herein encoding threonine synthase (SEQ ID NOs:16, 18, 20, 22, 24, and 26) and the *Arabidopsis thaliana* threonine synthase sequence set forth in GenBank Accession No. L41666 (SEQ ID NO:27).

- 5        Figure 5 shows a multiple alignment of the amino acid sequence fragments reported herein encoding threonine deaminase (SEQ ID NOs:29, 31, and 33) to the *Bruckholderia capacia* threonine synthase set forth in GenBank Accession No. U40630 (SEQ ID NO:34).

- Figure 6 shows the nucleotide sequence alignment of the S-adenosylmethionine synthetase reported herein for corn (SEQ ID NO:35) with the *Oryza sativa*  
10    S-adenosylmethionine synthetase nucleotide sequence set forth in EMBL Accession No. Z26867 (SEQ ID NO:37).

- Figure 7 shows the nucleotide sequence alignment of the S-adenosylmethionine synthetase reported here for soybean (SEQ ID NO:38) with the *Lycopersicon esculentum* S-adenosyl-methionine synthetase nucleotide sequence set forth in EMBL Accession  
15    No. Z24741 (SEQ ID NO:40).

- Figure 8 shows the nucleotide sequence alignment of the S-adenosylmethionine synthetase reported here for wheat (SEQ ID NO:41) with the *Hordeum vulgare* S-adenosylmethionine synthetase nucleotide sequence set forth in DDBJ Accession No. D63835 (SEQ ID NO:43).  
20    Amino acid sequence alignments were performed using the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153), from the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Nucleotide sequence alignments were a result of the BLASTN search performed with each individual S-adenosylmethionine sequence.

- 25    The following sequence descriptions and sequence listings attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

      SEQ ID NO:1 is the nucleotide sequence comprising the entire cDNA insert in clone csi1n.pk0042.a3 encoding a corn dihydrodipicolinate reductase.

- 30    SEQ ID NO:2 is the deduced amino acid sequence of a portion of a corn dihydrodipicolinate reductase derived from the nucleotide sequence of SEQ ID NO:1.

      SEQ ID NO:3 is the nucleotide sequence comprising a portion of the cDNA insert in clone rls2.pk0017.d3 encoding a rice dihydrodipicolinate reductase.

- SEQ ID NO:4 is the deduced amino acid sequence of a portion of a rice  
35    dihydrodipicolinate reductase derived from the nucleotide sequence of SEQ ID NO:3.

      SEQ ID NO:5 is the amino acid sequence of the entire *Synechocystis sp.* dihydrodipicolinate reductase DDBJ Accession No. D90899.

      SEQ ID NO:6 is the nucleotide sequence comprising the entire cDNA insert in clone chp2.pk0008.h4 encoding a corn diaminopimelate epimerase.

SEQ ID NO:7 is the deduced amino acid sequence of a portion of a corn diaminopimelate epimerase derived from the nucleotide sequence of SEQ ID NO:6.

SEQ ID NO:8 is the nucleotide sequence comprising a portion of the cDNA insert in clone rls48.pk0036.h10 encoding a rice diaminopimelate epimerase.

- 5        SEQ ID NO:9 is the deduced amino acid sequence of a portion of a rice diaminopimelate epimerase derived from the nucleotide sequence of SEQ ID NO:8.

SEQ ID NO:10 is the nucleotide sequence comprising a contig formed of portions of sfl1.pk0031.h3, and sgs1c.pk002.k12, and the entire cDNA insert from clones se2.pk0005.f1, and ses8w.pk0010.h11 encoding a soybean diaminopimelate epimerase.

- 10       SEQ ID NO:11 is the deduced amino acid sequence of a soybean diaminopimelate epimerase derived from the nucleotide sequence of SEQ ID NO:10.

SEQ ID NO:12 is the nucleotide sequence comprising a portion of the cDNA insert in clone wlm24.pk0030.g4 encoding a wheat diaminopimelate epimerase.

- 15       SEQ ID NO:13 is the deduced amino acid sequence of a portion of a wheat diaminopimelate epimerase derived from the nucleotide sequence of SEQ ID NO:12.

SEQ ID NO:14 is the nucleotide sequence comprising the entire *Synechocystis* sp. diaminopimelate epimerase DDBJ Accession No. D90917.

SEQ ID NO:15 is the nucleotide sequence comprising the entire cDNA insert in clone cc2.pk0031.c9 encoding a corn threonine synthase.

- 20       SEQ ID NO:16 is the deduced amino acid sequence of a portion of a corn threonine synthase derived from the nucleotide sequence set forth in SEQ ID NO:15.

SEQ ID NO:17 is the nucleotide sequence comprising part of the cDNA insert in clone cs1.pk0058.g5 encoding a corn threonine synthase.

- 25       SEQ ID NO:18 is the deduced amino acid sequence of a portion of a corn threonine synthase derived from the nucleotide sequence of SEQ ID NO:17.

SEQ ID NO:19 is the nucleotide sequence comprising part of the cDNA insert in clone rls72.pk0018.e7 encoding a rice threonine synthase.

SEQ ID NO:20 is deduced amino acid sequence of a portion of a rice threonine synthase derived from the nucleotide sequence set forth in SEQ ID NO:19.

- 30       SEQ ID NO:21 is the nucleotide sequence comprising part of the cDNA insert in clone sel.06a03 encoding a soybean threonine synthase.

SEQ ID NO:22 is the deduced amino acid sequence of a portion of a soybean threonine synthase derived from the nucleotide sequence set forth in SEQ ID NO:21.

- 35       SEQ ID NO:23 is the nucleotide sequence comprising the entire cDNA insert in clone srl.pk0003.f6 encoding a soybean threonine synthase.

SEQ ID NO:24 is the deduced amino acid sequence of a portion of a soybean threonine synthase derived from the nucleotide sequence set forth in SEQ ID NO:23.

SEQ ID NO:25 is the nucleotide sequence comprising part of the cDNA insert in clone wr1.pk0085.h2 encoding a wheat threonine synthase.

SEQ ID NO:26 is the deduced amino acid sequence of a portion of a wheat threonine synthase derived from the nucleotide sequence set forth in SEQ ID NO:25.

SEQ ID NO:27 is the entire amino acid sequence of an *Arabidopsis thaliana* threonine synthase found in GenBank Accession No. L41666.

5 SEQ ID NO:28 is the nucleotide sequence comprising the entire cDNA insert in clone cen1.pk0064.f4 encoding a corn threonine deaminase.

SEQ ID NO:29 is the deduced amino acid sequence of a portion of a corn threonine deaminase derived from the nucleotide sequence set forth in SEQ ID NO:28.

10 SEQ ID NO:30 is the nucleotide sequence comprising a portion of the cDNA insert in clone sfl1.pk0055.h7 encoding a soybean threonine deaminase.

SEQ ID NO:31 is the deduced amino acid sequence of a portion of a soybean threonine deaminase derived from the nucleotide sequence set forth in SEQ ID NO:30.

SEQ ID NO:32 is the nucleotide sequence comprising the entire cDNA insert in clone sre.pk0044.f3 encoding a soybean threonine deaminase.

15 SEQ ID NO:33 is the deduced amino acid sequence of a portion of a soybean threonine deaminase derived from the nucleotide sequence set forth in SEQ ID NO:32.

SEQ ID NO:34 is the entire amino acid sequence of a *Burkholderia capacia* threonine deaminase found in GenBank Accession No. U49630.

20 SEQ ID NO:35 is the nucleotide sequence comprising the entire cDNA insert in clone cc3.mn0002.d2 encoding the entire corn S-adenosylmethionine synthetase.

SEQ ID NO:36 is the deduced amino acid sequence of a corn S-adenosylmethionine synthetase derived from the nucleotide sequence set forth in SEQ ID NO:35.

SEQ ID NO:37 is the entire nucleotide sequence of a *Oryza sativa* S-adenosyl-methionine synthetase found in EMBL Accession No. Z26867.

25 SEQ ID NO:38 is the nucleotide sequence of the entire cDNA insert in clone s2.12b06 encoding the entire soybean S-adenosyl-methionine synthetase.

SEQ ID NO:39 is the deduced amino acid sequence of the entire soybean S-adenosyl-methionine synthetase derived from the nucleotide sequence set forth in SEQ ID NO:38.

30 SEQ ID NO:40 is the entire nucleotide sequence of a *Lycopersicon esculentum* S-adenosyl-methionine synthetase found in EMBL Accession No. Z24741.

SEQ ID NO:41 is the nucleotide sequence comprising a contig formed of portions of the cDNA inserts in clones wre1.pk0002.c12, wle1n.pk0070.b8, wkm1c.pk0003.g4, wlk1.pk0028.d3, wre1n.pk170.d8, wr1.pk0086.d5, wr1.pk0103.h8, and wre1n.pk0082.b2 encoding a portion of a wheat S-adenosyl-methionine synthetase.

35 SEQ ID NO:42 is the deduced amino acid sequence of a wheat S-adenosyl-methionine synthetase derived from the nucleotide sequence set forth in SEQ ID NO:41.

SEQ ID NO:43 is the entire nucleotide sequence of a *Hordeum vulgare* S-adenosyl-methionine synthetase found in DDBJ Accession No. D63835.

The Sequence Descriptions contain the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IYUB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

#### DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA. As used herein, "contig" refers to an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence. For example, several DNA sequences can be compared and aligned to identify common or overlapping regions. The individual sequences can then be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to

produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the sequences exemplified herein. Preferred substantially similar nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are 80% identical to the DNA sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are 90% identical to the identical to the DNA sequence of the nucleic acid fragments reported herein. Most preferred are nucleic acid fragments that are 95% identical to the DNA sequence of the nucleic acid fragments reported herein. The Clustal multiple alignment algorithm (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153) was used here with a GAP PENALTY of 10 and a GAP LENGTH PENALTY of 10.

A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.



“Codon degeneracy” refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the amino acid biosynthetic enzymes as set forth in SEQ ID NOs:2, 4, 7, 9, 11, 13, 16, 18, 20, 22, 24, 26, 29, 31, and 33. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

“Synthetic genes” can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. “Chemically synthesized”, as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

“Gene” refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature.

“Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a DNA sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the

associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The “translation leader sequence” refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G.D. (1995) *Molecular Biotechnology* 3:225).

The “3' non-coding sequences” refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell* 1:671-680.

“RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. “Messenger RNA (mRNA)” refers to the RNA that is without introns and that can be translated into protein by the cell. “cDNA” refers to a double-stranded DNA that is complementary to and derived from mRNA. “Sense” RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the

cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Pat. No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J.J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

“Transformation” refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as “transgenic” organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or “gene gun” transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Pat. No. 4,945,050).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter “Maniatis”).

Nucleic acid fragments encoding at least a portion of several plant amino acid biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the amino acid biosynthetic enzymes that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these enzymes.

**TABLE 1**  
Amino Acid Biosynthetic Enzymes

Enzyme	Clone	Plant
dihydrodipicolinate reductase	cs1.pk0083.b10	corn
	rls2.pk0017.d3	rice
diaminopimelate epimerase	chp2.pk0008.h4	corn
	rls48.pk0036.h10	rice
	se2.pk0005.f1	soybean
	ses8w.pk0010.f11	soybean
	sfl1.pk0031.h3	soybean
	sgs1c.pk002.k12	soybean
	wlm24.pk0030.g4	wheat
threonine synthase	cc2.pk0031.c9	corn
	cs1.pk0058.g5	corn
	rls72.pk0018.e7	rice
	se1.06a03	soybean
	sr1.pk0003.f6	soybean
	wr1.pk0085.h2	wheat
threonine deaminase	cen1.pk0064.f4	corn
	sfl1.pk0055.h7	soybean
	sre.pk0044.f3	soybean

Enzyme	Clone	Plant
s-adenosylmethionine synthase	cc3.mn0002.d2	corn
	se2.12b06	soybean
	wre1.pk0002.c12	wheat
	wle1n.pk0070.b8	wheat
	wkmlc.pk0003.g4	wheat
	wlk1.pk0028.d3	wheat
	wre1n.pk170.d8	wheat
	wr1.pk0086.d5	wheat
	wr1.pk0103.h8	wheat
	wre1n.pk0082.b2	wheat

The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous enzymes from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other amino acid biosynthetic enzymes, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primers DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part of or full-length of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow

the RACE protocol (Frohman et al., (1988) *PNAS USA* 85:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., (1989) *PNAS USA* 86:5673; Loh et al., (1989) *Science* 243:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M.A. and Martin, G.R., (1989) *Techniques* 1:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner, R.A. (1984) *Adv. Immunol.* 36:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed biosynthetic enzymes are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of free amino acids in those cells.

Overexpression of the biosynthetic enzymes of the instant invention may be accomplished by first constructing chimeric genes in which the coding region are operably linked to promoters capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric genes may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric genes may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric genes can then constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J.* 4:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant biosynthetic enzymes to different cellular compartments, or to facilitate their secretion from the cell. It is thus envisioned that the chimeric genes described above may be further supplemented by altering the coding sequences to encode enzymes with appropriate intracellular targeting sequences such as transit sequences (Keegstra, K. (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels, J.J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel, N. (1992) *Plant Phys.* 100:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of the genes encoding the instant biosynthetic enzymes in plants for some applications. In order to accomplish this, chimeric genes designed for co-suppression of the instant biosynthetic enzymes can be constructed by linking the genes or gene fragments encoding the enzymes to plant promoter sequences. Alternatively, chimeric genes designed to express antisense RNA for all or part of the instant nucleic acid fragments can be constructed by linking the genes or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant amino acid biosynthetic enzymes (or portions of the enzymes) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the enzymes by methods well known to those skilled in the art. The antibodies are useful for detecting the enzymes *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant amino acid biosynthetic enzymes are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct chimeric genes for production of the instant amino acid biosynthetic enzymes. These chimeric genes could then be introduced into appropriate microorganisms via transformation to provide high level expression of the enzymes. An example of a vector for high level expression of the instant amino acid biosynthetic enzymes in a bacterial host is provided (Example 11).

Additionally, the instant plant amino acid biosynthetic enzymes can be used as a targets to facilitate design and/or identification of inhibitors of the enzymes that may be useful as herbicides. This is desirable because the enzymes described herein catalyze various steps in a pathway leading to production of several essential amino acids. Accordingly, inhibition of the activity of one or more of the enzymes described herein could lead to inhibition of amino acid biosynthesis sufficient to inhibit plant growth. Thus, the instant plant amino acid biosynthetic enzymes could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al., (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein, D. et al., (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter* 4(1):37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel, J. D., et al., In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask, B. J. (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan, M. et al. (1995) *Genome Research* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med.* 114(2):95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics* 7:22-28) and Happy Mapping



(Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer, (1989) *Proc. Natl. Acad. Sci USA* 86:9402; Koes et al., (1995) *Proc. Natl. Acad. Sci USA* 92:8149; Bensen et al., (1995) *Plant Cell* 7:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase or S-adenosylmethionine synthetase. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding a dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase or S-adenosylmethionine synthetase can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the dihydrodipicolinate reductase, diaminopimelate epimerase, threonine synthase, threonine deaminase and S-adenosylmethionine synthetase gene product.

#### EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

5

TABLE 2  
cDNA Libraries from Corn and Soybean Tissues

Library	Tissue	Clone
cc2	Corn Callus, Partially Differentiated, 2 Weeks After Subculture	cc2.pk0031.c9
cc3	Corn Callus, Mature Somatic Embryo	cc3.mn0002.d2
cen1	Corn Endosperm 12 Days After Pollination	cen1.pk0064.f4
chp2	Corn Leaf, 11 Day Old Plant	chp2.pk0008.h4
cs1	Corn Leaf, Sheath 5 Week Old Plant	cs1.pk0058.g5
csi1n	Corn Silk*	csi1n.pk0042.a3
rls2	Rice Leaf 15 Days After Germination, 2 Hours After Infection of Strain Magaporthe grisea 4360-R-67 (AVR2-YAMO); Susceptible	rls2.pk0017.d3
rls48	Rice Leaf 15 Days After Germination, 48 Hours After Infection of Strain Magaporthe grisea 4360-R-67 (AVR2-YAMO); Susceptible	rls48.pk0036.h10
rls72	Rice Leaf 15 Days After Germination, 72 Hours After Infection of Strain Magaporthe grisea 4360-R-67 (AVR2-YAMO); Susceptible	rls72.pk0018.e7
s2	Soybean Seed, 19 Days After Flowering	s2.12b06
se1	Soybean Embryo 7 Days After Flowering	se1.06a03
se2	Soybean Embryo 10 Days After Flowering	se2.pk0005.f1
ses8w	Mature Soybean Embryo 8 Weeks After Subculture	ses8w.pk0010.h11
sfl1	Soybean Immature Flower	sfl1.pk0055.h7 sfl1.pk0031.h3
sgs1c	Soybean Seeds 4 Hours After Germination	sgs1c.pk002.k12
sr1	Soybean Root From 10 Day Old Seedlings	sr1.pk0003.f6
sre	Soybean Root Elongation 4-5 Days After Germination	sre.pk0044.f3
wkm1c	Wheat Kernel Malted 55 Hours at 22 Degrees Celsius	wkm1c.pk0003.g4
wle1n	Wheat Leaf From 7 Day Old Etiolated Seedling*	wle1n.pk0070.b8
wlk1	Wheat Seedlings 1 Hour After Treatment with Fungicide**	wlk1.pk0028.d3
wlm24	Wheat Seedlings 24 Hours After Inoculation With Erysiphe graminis f. sp tritici	wlm24.pk0030.g4
wr1	Wheat Root From 7 Day Old Seedling	wr1.pk0085.h2 wr1.pk0086.d5 wr1.pk0103.h8
wre1	Wheat Root From 7 Day Old Etiolated Seedling	wre1.pk0002.c12
wre1n	Wheat Root From 7 Day Old Etiolated Seedling*	wre1n.pk0082.b2 wre1n.pk170.d8

\*These libraries were normalized essentially as described in U.S. Pat. No. 5,482,845

\*\*Application of 6-iodo-2-propoxy-3-propyl-4(3*H*)-quinazolinone; synthesis and methods of using this compound are described in USSN 08/545,827, incorporated herein by reference.

- 5 cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing  
10 recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences, or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science* 252:1651).  
15 The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

#### EXAMPLE 2

##### Identification and Characterization of cDNA Clones

- ESTs encoding plant amino acid biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.*  
20 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity  
25 to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics* 3:266-272) provided by the  
30 NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

35

#### EXAMPLE 3

##### Characterization of cDNA Clones Encoding Polypeptides Homologous to Dihydrodipicolinate Reductase

The BLASTX search using the nucleotide sequences from clones csi1n.pk0042.a3 and rls2.pk0017.d3 revealed similarity of the protein encoded by the cDNA to *Synechocystis sp.*

dihydrodipicolinate reductase enzyme (DDBJ Accession No. D90899). BLAST pLog values were 12.60 and 11.68 for csi1n.pk0042.a3 and rls2.pk0017.d3, respectively.

The sequence of the entire cDNA insert in clone csi1n.pk0042.a3 was determined and is shown in SEQ ID NO:1; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:2. The amino acid sequence set forth in SEQ ID NO:2 was evaluated by BLASTP, yielding a pLog value of 36.72 versus the *Synechocystis sp.* dihydrodipicolinate reductase sequence. The sequence of a portion of the cDNA insert from clone rls2.pk0017.d3 is shown in SEQ ID NO:3; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:4. Figure 2 presents an alignment of the amino acid sequences set forth in SEQ ID NO:2 and the *Synechocystis sp.* dihydrodipicolinate reductase sequence (SEQ ID NO:5). SEQ ID NO:2 is 40% identical to the *Synechocystis sp.* dihydrodipicolinate reductase sequence (SEQ ID NO:5). Sequence alignments were performed by the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153), using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Sequence percent identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz.* 183:626-645) using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode a nearly entire corn dihydrodipicolinate reductase, and a portion of a rice dihydrodipicolinate reductase. These sequences represent the first plant sequences encoding dihydrodipicolinate reductase.

#### EXAMPLE 4

##### Characterization of cDNA Clones Encoding Diaminopimelate Epimerase

The BLASTX search using the nucleotide sequences from clones chp2.pk0008.h4, rls48.pk0036.h10, wlm24.pk0030.g4, and the contig sequences assembled from clones se2.pk0005.fl, ses8w.pk0010.h11, sfl1.pk0031.h3, and sgs1c.pk0002.k12 revealed similarity of the proteins encoded by the cDNAs to diaminopimelate epimerase from *Synechocystis sp.* (DDBJ Accession No. D90917). The BLAST results for each of these ESTs are shown in Table 3:

30

TABLE 3

BLAST Results for Clones Encoding Polypeptides Homologous  
to Diaminopimelate Epimerase

Clone	BLAST pLog Score DDBJ Accession No. D90917
chp2.pk0008.h4	59.16
rls48.pk0036.h10	40.82
The contig of:	98.30
se2.pk0005.f1	
ses8w.pk0010.h11	
sfl1.pk0031.h3	
sgs1c.pk002.k12	
wlm24.pk0030.g4	23.46

- 5       The sequence of the entire cDNA insert in clone chp2.pk0008.h4 was determined and is shown in SEQ ID NO:6; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:7. The amino acid sequence set forth in SEQ ID NO:7 was evaluated by BLASTP, yielding a pLog value of 75.66 versus the *Synechocystis sp.* sequence. The sequence of a portion of the cDNA insert from clone rls48.pk0036.h10 is shown in SEQ ID NO:8; the
- 10       deduced amino acid sequence of this cDNA is shown in SEQ ID NO:9. The nucleotide sequence of the contig assembled from clones se2.pk0005.f1, ses8w.pk0010.h11, sfl1.pk0031.h3, and sgs1c.pk002.k12 was determined and is shown in SEQ ID NO:10; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:11. The amino acid sequence set forth in SEQ ID NO:11 was evaluated by BLASTP, yielding a pLog value of
- 15       98.57 versus the *Synechocystis sp.* sequence. The sequence of a portion of the cDNA insert from clone wlm24.pk0030.g4 is shown in SEQ ID NO:12; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:13. Figure 3 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:7, 9, 11, and 13 and the *Synechocystis sp.* sequence (SEQ ID NO:14). The data in Table 4 represents a calculation of the percent identity of the
- 20       amino acid sequences set forth in SEQ ID NOs: 7, 9, 11, and 13 and the *Synechocystis sp.* sequence.

**TABLE 4**

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of  
cDNA Clones Encoding Polypeptides Homologous  
to Diaminopimelate Epimerase

Clone	SEQ ID NO.	Percent Identity to DDBJ Accession No. D90917 (SEQ ID NO:16)
chp2.pk0008.h4	7	59
rls48.pk0036.h10	9	74
Contig of:	11	72
se2.pk0005.fl		
ses8w.pk0010.h11		
sfl1.pk0031.h3		
sgs1c.pk002.k12		
wlm24.pk0030.g4	13	65

5 Sequence alignments were performed by the Clustal method of alignment (Higgins,  
D.G. and Sharp, P.M. (1989) CABIOS 5:151-153), using the Megalign program of the  
LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Sequence  
percent identity calculations were performed by the Jotun Hein method (Hein. J. J. (1990)  
10 Meth. Enz. 183:626-645) using the Megalign program of the LASARGENE bioinformatics  
computing suite (DNASTAR Inc., Madison, WI).

Sequence alignments and BLAST scores and probabilities indicate that the instant  
nucleic acid fragments encode a nearly entire corn diaminopimelate epimerase  
(chp2.pk0008.h4), a portion of a rice diaminopimelate epimerase (rls48.pk0036.h10), and an  
15 entire soybean diaminopimelate epimerase (se2.pk0005.fl, ses8w.pk0010.h11,  
sfl1.pk0031.h3, and sgs1c.pk002.k12), and a portion of a wheat diaminopimelate epimerase  
(wlm24.pk0030.g4). These sequences represent the first plant sequences encoding  
diaminopimelate epimerase enzyme.

**EXAMPLE 5****20 Characterization of cDNA Clones Encoding Threonine Synthase**

The BLASTX search using the EST sequences from clones cc2.pk0031.c9,  
cs1.pk0058.g5, rls72.pk0018.e7, se1.06a03, srl.pk0003.f6, and wr1.pk0085.h2 revealed  
similarity of the proteins encoded by the cDNAs to threonine synthase from *Arabidopsis*  
*thaliana* (GenBank Accession No. L41666). The BLAST results for each of these ESTs are  
25 shown in Table 5:

TABLE 5  
BLAST Results for Clones Encoding Polypeptides Homologous  
to Threonine Synthase

Clone	BLAST pLog Score
	L41666
cc2.pk0031.c9	56.19
cs1.pk0058.g5	8.00
rls72.pk0018.e7	29.47
se1.06a03	34.15
sr1.pk0003.f6	21.13
wr1.pk0085.h2	29.47

- 5        The sequence of the entire cDNA insert in clone cc2.pk0031.c9 was determined and is shown in SEQ ID NO:15; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:16. The amino acid sequence set forth in SEQ ID NO:16 was evaluated by BLASTP, yielding a pLog value of 166.11 versus the *Arabidopsis thaliana* sequence. BLASTN
- 10        against dbest indicated identity of nucleotides 520 through 684 from cc2.pk0031.c9 with nucleotides 1 through 162 of a corn EST (GenBank Accession No. T18847). The sequence of a portion of the cDNA insert from clone cs1.pk0058.g5 is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:18. The sequence of a portion of the cDNA insert from clone rls72.pk0018.e7 is shown in SEQ ID NO:19; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:20. The sequence of a
- 15        portion of the cDNA insert from clone se1.06a03 is shown in SEQ ID NO:21; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:22. The sequence of the entire cDNA insert in clone sr1.pk0003.f6 was determined and is shown in SEQ ID NO:23; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:24. The amino acid sequence set forth in SEQ ID NO:24 was evaluated by BLASTP, yielding a pLog value of
- 20        275.06 versus the *Arabidopsis thaliana* sequence. The sequence of a portion of the cDNA insert from clone wr1.pk0085.h2 is shown in SEQ ID NO:25; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:26. Figure 4 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:16, 18, 20, 22, 24, and 26 and the *Arabidopsis thaliana* sequence. The data in Table 6 represents a calculation of the percent
- 25        identity of the amino acid sequences set forth in SEQ ID NOs:16, 18, 20, 22, 24, and 26 and the *Arabidopsis thaliana* sequence (SEQ ID NO:27).

TABLE 6

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of  
cDNA Clones Encoding Polypeptides Homologous  
to Threonine Synthase

Clone	SEQ ID NO.	Percent Identity to L41666 (SEQ ID NO:29)
cc2.pk0031.c9	16	81.0
cs1.pk0058.g5	18	81.0
rls72.pk0018.e7	20	55.3
se1.06a03	22	80.0
sr1.pk0003.f6	24	84.4
wr1.pk0085.h2	26	50.4

5

Sequence alignments were performed by the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153), using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Sequence percent identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz.* 183:626-645) using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI).

10

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode portions of a corn threonine synthase (cc2.pk0031.c9 and cs1.pk0058.g5), a portion of a rice threonine synthase (rls72.pk0018.e7), portions of a soybean threonine synthase (se1.06a03 and sr1.pk0003.f6), and a portion of a wheat threonine synthase (wr1.pk0085.h2). These sequences represent the first corn, rice, soybean, and wheat sequences encoding threonine synthase.

15

#### EXAMPLE 6

##### Characterization of cDNA Clones Encoding Threonine Deaminase

20

The BLASTX search using the EST sequence from clone cen1.pk0064.f4 revealed similarity of the protein encoded by the cDNA to threonine deaminase from *Brukholderia capacia* (GenBank Accession No. U40630; pLog = 31.38). The BLASTX search using the EST sequences from clones sf11.pk0055.h7 and sre.pk0044.f3 revealed similarity of the proteins encoded by the cDNAs to threonine deaminase from *Solanum tuberosum* and *Brukholderia capacia* (EMBL Accession No. X67846 and GenBank Accession No. U40630, respectively). BLAST pLog values were 36.55 and 31.79 for sf11.pk0055.h7, and 19.47 and 14.51 for sre.pk0044.f3.

25

The sequence of the entire cDNA insert in clone cen1.pk0064.f4 was determined and is shown in SEQ ID NO:28; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:29. The amino acid sequence set forth in SEQ ID NO:29 was evaluated by BLASTP, yielding a pLog value of 134.85 versus the *Brukholderia capacia* sequence. The sequence of a portion of the cDNA insert from clone sf11.pk0055.h7 is shown in SEQ ID

30



NO:30; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:31. The sequence of the entire cDNA insert in clone sre.pk0044.f3 was determined and is shown in SEQ ID NO:32; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:33. The amino acid sequence set forth in SEQ ID NO:33 was evaluated by BLASTP, yielding pLog values of 19.24 versus the *Solanum tuberosum* sequence and 15.19 versus the *Brukholderia capacia* threonine deaminase sequence. Figure 5 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:29, 31, and 33 and the *Brukholderia capacia* (SEQ IDNO:34) sequence. The data in Table 7 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:29, 31, and 33 and the *Brukholderia capacia* sequence.

TABLE 7

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Threonine Deaminase

Clone	SEQ ID NO.	Percent Identity to U40630 (SEQ ID NO:36)
cen1.pk0064.f4	29	61.0
sfl1.pk0055.h7	31	47.9
sre.pk0044.f3	33	46.0

Sequence alignments were performed by the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153), using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Sequence percent identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz.* 183:626-645) using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI)

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode a nearly entire corn threonine deaminase (cen1.pk0064.f4) and portions of a soybean threonine deaminase (sfl1.pk0055.h7 and sre.pk0044.f3). These sequences represent the first corn and soybean sequences encoding threonine deaminase.

## EXAMPLE 7

Characterization of cDNA Clones Encoding S-adenosylmethionine synthetase

The BLASTX search using the nucleotide sequence from clone cc3.mn0002.d2 revealed similarity of the protein encoded by the cDNA to S-adenosylmethionine synthetase from *Oryza sativa* (EMBL Accession No. Z26867; pLog = 99.03). The sequence of the entire cDNA insert in clone cc3.mn0002.d2 was determined and is shown in SEQ ID NO:35; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:36. The nucleotide sequence set forth in SEQ ID NO:35 was evaluated by BLASTN, yielding a pLog value larger than 200 versus the *Oryza sativa* sequence. Figure 6 presents an alignment of the

nucleotide sequences set forth in SEQ ID NO:35 and the *Oryza sativa* sequence (SEQ ID NO:37). The nucleotide sequence in SEQ ID NO:35 is 88% identical over 1216 nucleotides to the nucleotide sequence of the *Oryza sativa* S-adenosylmethionine synthetase.

5 The BLASTX search using the nucleotide sequence from clone s2.12b06 revealed similarity of the protein encoded by the cDNA to S-adenosylmethionine synthetase from *Lycopersicon esculentum* (EMBL Accession No. Z24741; pLog = 62.62). The sequence of the entire cDNA insert in clone s2.12b06 was determined and is shown in SEQ ID NO:38; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:39. The nucleotide sequence set forth in SEQ ID NO:38 was evaluated by BLASTN, yielding a pLog value  
10 larger than 200 versus the *Lycopersicon esculentum* sequence. Figure 7 presents an alignment of the nucleotide sequences set forth in SEQ ID NO:38 and the *Lycopersicon esculentum* sequence (SEQ ID NO:40). The nucleotide sequence set forth in SEQ ID NO:38 is 82 % identical over 1210 nucleotides to the *Lycopersicon esculentum* sequence.

The BLASTX search using the nucleotide sequence from the contig assembled from  
15 clones wre1.pk0002.c12, wle1n.pk0070.b8, wkm1c.pk0003.g4, wlkl.pk0028.d3, wre1n.pk170.d8, wr1.pk0086.d5, wr1.pk0103.h8, and wre1n.pk0082.b2 revealed similarity of the protein encoded by the contig to S-adenosylmethionine synthetase from *Hordeum vulgare* (DDBJ Accession No. D63835) with a pLog value larger than 200. The nucleotide sequence of the contig assembled from clones wre1.pk0002.c12, wle1n.pk0070.b8,  
20 wkm1c.pk0003.g4, wlkl.pk0028.d3, wre1n.pk170.d8, wr1.pk0086.d5, wr1.pk0103.h8, and wre1n.pk0082.b2 is shown in SEQ ID NO:41; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:42. Figure 8 presents an alignment of the nucleotide sequence set forth in SEQ ID NO:41 and the *Hordeum vulgare* sequence (SEQ ID NO:43). The SEQ ID NO:41 is 92% identical to the *Hordeum vulgare* sequence.

25 Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire or nearly entire corn, soybean, or wheat S-adenosylmethionine synthetase. These sequences represent the first corn, soybean, or wheat sequences encoding S-adenosylmethionine synthetase.

#### EXAMPLE 8

##### Expression of Chimeric Genes in Monocot Cells

30 A chimeric gene comprising a cDNA encoding an amino acid biosynthetic enzyme in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain  
35 reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers and under appropriate experimental conditions. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. The amplified DNA can then be digested with restriction enzymes NcoI and SmaI and fractionated on a 0.7% low melting point

agarose gel in 40 mM Tris-acetate, pH 8.5, 1 mM EDTA. The appropriate band can be excised from the gel, melted at 68°C and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Boulevard,  
5 Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli*  
10 XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment  
15 encoding a plant amino acid biosynthetic enzyme, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed  
20 with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can  
25 be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers  
30 resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al., (1987) *Nature* 327:70-73) may be  
35 used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions.

After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200  $\mu$ L of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30  $\mu$ L of ethanol. An aliquot (5  $\mu$ L) of the DNA-coated  
5 gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-  
10 solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock  
15 tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains glufosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter  
20 of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the  
25 tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology* 8:833-839).

### EXAMPLE 9

#### Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription  
30 terminator from the gene encoding the  $\beta$  subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant amino acid biosynthetic enzymes in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of  
35 phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be

incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Plant amino acid biosynthetic enzymes are known to be localized in the chloroplasts. Accordingly, for those enzymes (or polypeptides representing part of the instant amino acid biosynthetic enzymes) that lack a chloroplast targeting signal, the DNA fragment to be inserted into the expression vector can be synthesized by PCR with primers encoding a chloroplast targeting signal. For example, a chloroplast transit sequence equivalent to the cts of the small subunit of ribulose 1,5-bisphosphate carboxylase from soybean (Berry-Lowe et al. (1982) *J. Mol. Appl. Genet.* 1:483-498) may be used.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding a plant amino acid biosynthetic enzyme. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with fluorescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Kline et al. (1987) *Nature* (London) 327:70, U.S. Patent No. 4,945,050). A Du Pont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the biosynthetic enzyme and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µL spermidine (0.1 M), and 50 µL CaCl<sub>2</sub> (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and

resuspended in 40  $\mu$ L of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five  $\mu$ L of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

#### EXAMPLE 10

##### Analysis of Amino Acid Content of the Seeds of Transformed Plants

To analyze for expression of the chimeric genes in seeds and for the consequences of expression on the amino acid content in the seeds, a seed meal can be prepared by any of a number of suitable methods known to those skilled in the art. The seed meal can be partially or completely defatted, via hexane extraction for example, if desired. Protein extracts can be prepared from the meal and analyzed for enzyme activity. Alternatively the presence of any of the expressed enzymes can be tested for immunologically by methods well-known to those skilled in the art. To measure free amino acid composition of the seeds, free amino acids can be extracted from the meal and analyzed by methods known to those skilled in the art (Bielecki et al. (1966) *Anal. Biochem.* 17:278-293). Amino acid composition can then be determined using any commercially available amino acid analyzer. To measure total amino acid composition of the seeds, meal containing both protein-bound and free amino acids can be acid hydrolyzed to release the protein-bound amino acids and the composition can then be determined using any commercially available amino acid analyzer. Seeds expressing the instant amino acid biosynthetic enzymes and with altered lysine, threonine, methionine, cysteine and/or isoleucine content as compared to the wild type seeds can thus be identified and propagated.

To measure free amino acid composition of the seeds, free amino acids can be extracted from 8-10 milligrams of the seed meal in 1.0 mL of methanol/chloroform/water

mixed in ratio of 12v/5v/3v (MCW) at room temperature. The mixture can be vortexed and then centrifuged in an eppendorf microcentrifuge for about 3 min; approximately 0.8 mL of supernatant is then decanted. To this supernatant, 0.2 mL of chloroform is added followed by 0.3 mL of water. The mixture is then vortexed and centrifuged in an eppendorf microcentrifuge for about 3 min. The upper aqueous phase, approximately 1.0 mL, can then be removed and dried down in a Savant Speed Vac Concentrator. The samples are then hydrolyzed in 6N hydrochloric acid, 0.4%  $\beta$ -mercaptoethanol under nitrogen for 24 h at 110-120°C. Ten percent of the sample can then be analyzed using a Beckman Model 6300 amino acid analyzer using post-column ninhydrin detection.

Relative free amino acid levels in the seeds are then compared as ratios of lysine, threonine, methionine, cysteine and/or isoleucine to leucine, thus using leucine as an internal standard.

### EXAMPLE 11

#### Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant plant amino acid biosynthetic enzymes can be inserted into the T7 *E. coli* expression vector pET24d (Novagen). Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the enzyme. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10  $\mu$ g/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20  $\mu$ L of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pET24d is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pET24d and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing 2xYT media and 50  $\mu$ g/mL kanamycin.

Transformants containing gene encoding the enzyme are then screened for the correct orientation with respect to pET24d T7 promoter by restriction enzyme analysis.

Clones in the correct orientation with respect to the T7 promoter can be transformed into BL21(DE3) competent cells (Novagen) and selected on 2xYT agar plates containing 50  $\mu$ g/ml kanamycin. A colony arising from this transformation construct can be grown overnight at 30°C in 2xYT media with 50  $\mu$ g/mL kanamycin. The culture is then diluted two fold with fresh media, allowed to re-grow for 1 h, and induced by adding isopropyl-thiogalactopyranoside to 1 mM final concentration. Cells are then harvested by centrifugation after 3 h and re-suspended in 50  $\mu$ L of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass

beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One  $\mu\text{g}$  of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

#### EXAMPLE 12

##### Evaluating Compounds for Their Ability to Inhibit the Activity of a Plant Amino Acid Biosynthetic Enzyme

The plant amino acid biosynthetic enzymes described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 6, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant enzymes may be expressed separately as mature proteins, or may be co-expressed in *E. coli* or another suitable expression background. In addition, whether expressed separately or in combination, the instant enzymes may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("His)<sub>6</sub>"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzymes. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the biosynthetic enzyme.

Purification of the instant enzymes, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the enzymes are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, an enzyme may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a (His)<sub>6</sub> peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include  $\beta$ -mercaptoethanol or other



reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the biosynthetic enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

- 5        Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the plant amino acid biosynthetic enzymes disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. Examples of assays for many of these enzymes can be found in *Methods in*
- 10    *Enzymology* Vol. V, (Colowick and Kaplan eds.) Academic Press, New York or *Methods in Enzymology* Vol. XVII, (Tabor and Tabor eds.) Academic Press, New York. Specific examples may be found in the following references, each of which is incorporated herein by reference: dihydrodipicolinate reductase may be assayed as described in Farkas et al. (1965) *J. Biol. Chem.* 240: 4717-4722, or Cremer et al. (1988) *J. Gen. Microbiol.* 134:3221-3229;
- 15    diaminopimelate epimerase may be assayed as described in Work (1962) in *Methods in Enzymology* Vol. V, (Colowick and Kaplan eds.) 858-864, Academic Press, New York; threonine synthase may be assayed as described in Giovanelli et al. (1984) *Plant Physiol* 76: 285-292 or Curien et al. (1996) *FEBS Lett.* 390: 85-90; threonine deaminase may be assayed as described in Tomova et al. (1968) *Biochemistry (USSR)* 33: 200-208 or Dougall (1970)
- 20    *Phytochemistry* 9: 959-964; and S-adenosylmethionine synthetase may be assayed as described in Mudd (1960) *Biochim. Biophys. Acta* 38:354-355 or Boerjan et al. (1994) *Plant Cell* 6:1401-1414.